

PATENT ABSTRACTS OF JAPAN

(11) Publication number : 2002-311979
 (43) Date of publication of application : 25.10.2002

(51) Int.CI.

G10L 13/06
 G10L 13/00
 G10L 13/08

(21) Application number : 2001-118417

(71) Applicant : SONY CORP

(22) Date of filing : 17.04.2001

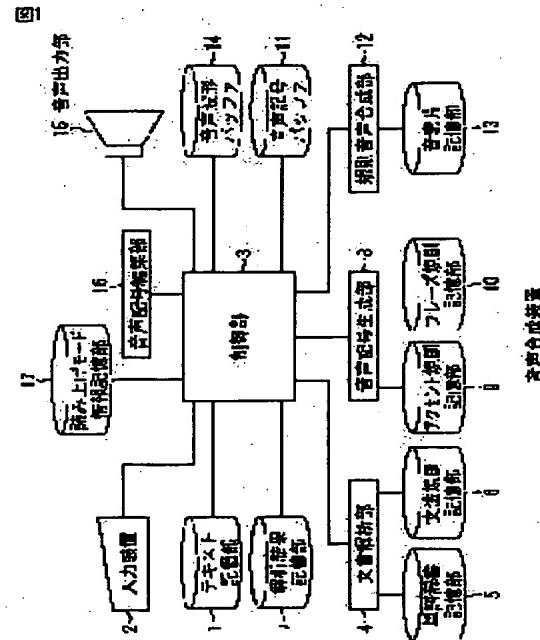
(72) Inventor : KOBAYASHI KENICHIRO
 YAMAZAKI NOBUHIDE
 ASANO KOJI
 KARIYA SHINICHI
 FUJITA YAEKO

(54) SPEECH SYNTHESIZER, SPEECH SYNTHESIS METHOD, PROGRAM AND RECORDING MEDIUM

(57) Abstract:

PROBLEM TO BE SOLVED: To create a synthetic sound with a natural stammer.

SOLUTION: A speech sign editing part 16 converts a speech sign into a speech sign in the state where a prescribed phoneme in the speech sign is repeated by editing the speech sign according to read-aloud mode information. The speech sign editing part 16 further carries out editing to control the speech speed and power of the phoneme portion to be repeated with respect to the speech sign according to read-aloud mode information. Then, a rules speech synthesis part 12 carries out a ruled speech synthesis according to the speech sign after editing.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision
of rejection]

[Date of requesting appeal against examiner's
decision of rejection]

[Date of extinction of right]

(19) 日本国特許庁 (J P)

(12) 公 開 特 許 公 報 (A)

(11)特許出願公開番号

特開2002-311979

(P2002-311979A)

(43) 公開日 平成14年10月25日(2002.10.25)

(51) Int.Cl.⁷
G 1 0 L 13/06
13/00
13/08

識別記号

F I
C 1 O L 5/04
3/00

テーマコード(参考)
5D046

審査請求 未請求 請求項の数23 OL (全 11 頁)

(21)出願番号 特願2001-118417(P2001-118417)

(71)出願人 000002185
ソニー株式会社
東京都品川区北品川6丁目7番35号

(72)発明者 小林 貴一郎
東京都品川区北品川6丁目7番35号 ソニ
一株式会社内

(72)発明者 山崎 信英
東京都品川区北品川6丁目7番35号 ソニ
一株式会社内

(74)代理人 100082131
弁理士 稲本 義雄

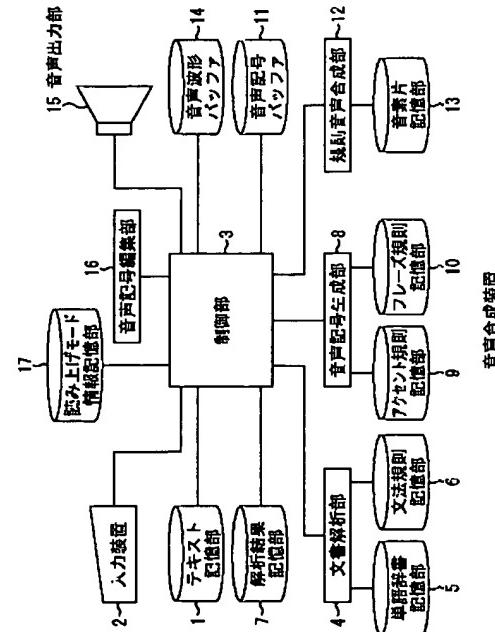
(54) 【発明の名称】 音声合成装置および音声合成方法、並びにプログラムおよび記録媒体

(57) 【要約】

【課題】 自然な吃りのある合成音を生成する。

【解決手段】 音声記号編集部16は、音声記号を、読み上げモード情報にしたがって編集することにより、音声記号における所定の音韻を反復させた状態の音声記号に変換する。さらに、音声記号編集部16は、読み上げモード情報にしたがい、音声記号について、反復させる音韻部分の発話速度やパワーを制御する編集も行う。そして、規則音声合成部12は、その編集後の音声記号にしたがって規則音声合成を行う。

四



【特許請求の範囲】

【請求項1】 少なくとも音韻情報を含む音声情報から、合成音を生成する音声合成装置であって、前記音声情報を、所定の規則にしたがって編集することにより、前記音声情報における所定の音韻を反復させた状態の音声情報に変換する編集手段と、前記編集手段において編集された前記音声情報にしたがい、合成音を生成する音声合成手段とを備えることを特徴とする音声合成装置。

【請求項2】 前記編集手段は、前記音声情報における文頭の音韻を反復させることを特徴とする請求項1に記載の音声合成装置。

【請求項3】 前記編集手段は、前記音声情報がポーズを表すポーズ記号を含む場合に、そのポーズ記号の直後の音韻を反復させることを特徴とする請求項1に記載の音声合成装置。

【請求項4】 前記編集手段は、前記音声情報が、前記編集手段に対して、音韻を反復させる編集を行うことを指示する編集用タグを含む場合に、その編集用タグにしたがって、音韻を反復させることを特徴とする請求項1に記載の音声合成装置。

【請求項5】 前記編集手段は、前記所定の音韻を、一定回数だけ反復させることを特徴とする請求項1に記載の音声合成装置。

【請求項6】 前記編集手段は、前記所定の音韻を、乱数により決定される回数だけ反復させることを特徴とする請求項1に記載の音声合成装置。

【請求項7】 前記乱数により決定される回数について、その上限または下限が設定可能となっていることを特徴とする請求項6に記載の音声合成装置。

【請求項8】 前記乱数により決定される回数の上限または下限を入力する入力手段をさらに備えることを特徴とする請求項7に記載の音声合成装置。

【請求項9】 前記編集手段は、前記編集用タグにおいて、音韻の反復回数が指示されている場合に、その反復回数だけ、音韻を反復させることを特徴とする請求項4に記載の音声合成装置。

【請求項10】 前記編集手段は、音韻を反復させるとともに、その反復される音韻部分の発話速度が他の部分よりも速くなるように、前記音声情報を編集することを特徴とする請求項1に記載の音声合成装置。

【請求項11】 前記編集手段は、反復される音韻部分の発話速度を、乱数によって決定することを特徴とする請求項10に記載の音声合成装置。

【請求項12】 前記乱数により決定される発話速度について、その上限または下限が設定可能となっていることを特徴とする請求項11に記載の音声合成装置。

【請求項13】 前記編集手段は、反復される音韻部分の発話速度を、その音韻の反復回数に基づいて決定することを特徴とする請求項10に記載の音声合成装置。

【請求項14】 前記編集手段は、反復される音韻部分の発話速度が変化するように、前記音声情報を編集することを特徴とする請求項10に記載の音声合成装置。

【請求項15】 前記編集手段は、音韻を反復させるとともに、その反復される音韻部分の直後の音韻のパワーが大きくなるように、前記音声情報を編集することを特徴とする請求項1に記載の音声合成装置。

【請求項16】 前記編集手段は、音韻を反復させるとともに、その反復される音韻部分のパワーが変化するよう、前記音声情報を編集することを特徴とする請求項1に記載の音声合成装置。

【請求項17】 前記編集手段は、音韻を反復させるよう、前記音声情報を編集するとともに、前記音声合成手段に対して、合成音の発話速度またはパワーを指示する合成用タグを、前記音声情報に挿入することを特徴とする請求項1に記載の音声合成装置。

【請求項18】 前記編集手段は、前記音声情報が、前記編集手段に対して、合成音の発話速度またはパワーを制御する編集を行うことを指示する編集用タグを含む場合に、その編集用タグにしたがって、前記音声情報を、さらに編集することを特徴とする請求項1に記載の音声合成装置。

【請求項19】 前記編集手段は、前記音声情報における音韻情報を繰り返し記述することにより、その音韻情報に対応する所定の音韻を反復させた状態の音声情報を得ることを特徴とする請求項1に記載の音声合成装置。

【請求項20】 テキストを言語処理する言語処理手段と、

前記言語処理手段による言語処理結果に基づいて、前記音声情報を生成する生成手段とをさらに備えることを特徴とする請求項1に記載の音声合成装置。

【請求項21】 少なくとも音韻情報を含む音声情報から、合成音を生成する音声合成方法であって、前記音声情報を、所定の規則にしたがって編集することにより、前記音声情報における所定の音韻を反復させた状態の音声情報に変換する編集ステップと、前記編集ステップにおいて編集された前記音声情報にしたがい、合成音を生成する音声合成ステップとを備えることを特徴とする音声合成方法。

【請求項22】 少なくとも音韻情報を含む音声情報から、合成音を生成する音声合成処理を、コンピュータに行わせるプログラムであって、

前記音声情報を、所定の規則にしたがって編集することにより、前記音声情報における所定の音韻を反復させた状態の音声情報に変換する編集ステップと、

前記編集ステップにおいて編集された前記音声情報にしたがい、合成音を生成する音声合成ステップとを備えることを特徴とするプログラム。

【請求項23】 少なくとも音韻情報を含む音声情報から、合成音を生成する音声合成処理を、コンピュータに

行わせるプログラムが記録されている記録媒体であつて、前記音声情報を、所定の規則にしたがって編集することにより、前記音声情報における所定の音韻を反復させた状態の音声情報に変換する編集ステップと、前記編集ステップにおいて編集された前記音声情報にしたがい、合成音を生成する音声合成ステップとを備えるプログラムが記録されていることを特徴とする記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、音声合成装置および音声合成方法、並びにプログラムおよび記録媒体に関するものである。

{(　　/　　)P3(　　)}

【0005】音声記号(1)において、カタカナは、音韻情報を表している。また、「{}」、「()」、「'」、「/」、「P3」は、タグであり、タグ'は、アクセントを表す。さらに、タグ{}および()は、フレーズの区切りを表し、タグP3のうちのPは、ポーズを表し、続く数字の3は、ポーズの長さを表す。また、タグ/は、アクセント句の区切りを表す。なお、これらのタグは、音声合成に用いられるので、以下、適宜、合成用タグという。

【0006】音声記号(1)によれば、音声合成装置では、音韻が「これまで進められた研究は、大きな成果を……」という合成音が生成される。

【0007】

【発明が解決しようとする課題】ところで、最近、例えば、ペット型のペットロボット等として、音声合成装置を搭載し、ユーザーに話しかけたり、ユーザーと会話(対話)を行うものが提案されており、このようなペットロボットの発話(合成音)に、個性を持たせることができれば、よりエンタテイメント性を向上させることができるものである。

【0008】合成音に個性を持たせる方法としては、合成音の発話速度や、高さ、強さ等を制御する方法が考えられるが、その他、例えば、吃りを実現することができれば、合成音に、より個性を持たせることができると考えられる。

【0009】合成音を吃らせる方法としては、例えば、音声合成装置に入力するテキストを構成する文字を、反復して記述する方法があるが、テキストを、その中の文字を反復して記述するのは面倒である。さらに、単に、文字を反復したテキストを入力しても、自然な感じの吃りがある合成音を得ることは困難である。

【0010】本発明は、このような状況に鑑みてなされたものであり、吃りのある合成音を、容易に生成することができるようとするものである。

【0011】

【課題を解決するための手段】本発明の音声合成装置

し、特に、吃りのある合成音を容易に生成することができるようとする音声合成装置および音声合成方法、並びにプログラムおよび記録媒体に関するものである。

【0002】

【従来の技術】従来の音声合成装置においては、テキスト、またはそのテキストを解析して得られる音声記号に基づいて、合成音が生成される。また、音声合成装置では、例えば、合成音の発話速度や、高さ、強さ等が、音声合成装置に入力されるテキストや音声記号(列)に挿入されたタグに基づいて制御される。

【0003】ここで、タグが挿入された音声記号としては、例えば、次のようなものがある。

【0004】

/') . . . }

. . . (1)

は、音声情報を、所定の規則にしたがって編集することにより、音声情報における所定の音韻を反復させた状態の音声情報に変換する編集手段と、編集手段において編集された音声情報にしたがい、合成音を生成する音声合成手段とを備えることを特徴とする。

【0012】本発明の音声合成方法は、音声情報を、所定の規則にしたがって編集することにより、音声情報における所定の音韻を反復させた状態の音声情報に変換する編集ステップと、編集ステップにおいて編集された音声情報にしたがい、合成音を生成する音声合成ステップとを備えることを特徴とする。

【0013】本発明のプログラムは、音声情報を、所定の規則にしたがって編集することにより、音声情報における所定の音韻を反復させた状態の音声情報に変換する編集ステップと、編集ステップにおいて編集された音声情報にしたがい、合成音を生成する音声合成ステップとを備えることを特徴とする。

【0014】本発明の記録媒体は、音声情報を、所定の規則にしたがって編集することにより、音声情報における所定の音韻を反復させた状態の音声情報に変換する編集ステップと、編集ステップにおいて編集された音声情報にしたがい、合成音を生成する音声合成ステップとを備えるプログラムが記録されていることを特徴とする。

【0015】本発明の音声合成装置および音声合成方法、並びにプログラムにおいては、音声情報が、所定の規則にしたがって編集され、音声情報における所定の音韻を反復させた状態の音声情報に変換される。そして、その編集された音声情報にしたがい、合成音が生成される。

【0016】

【発明の実施の形態】図1は、本発明を適用した音声合成装置の一実施の形態の構成例を示している。

【0017】テキスト記憶部1は、この音声合成装置によって、合成音による読み上げを行う文書を、例えば、テキストデータの形で記憶している。このテキスト記憶

部1に記憶されたテキストデータ(文書データ)は、制御部3によって読み出されるようになっている。

【0018】入力装置2は、キーボードや、ポインティングデバイス、あるいは、マイクロフォンと音声認識装置等で構成されており、ユーザが所定の指示等を入力するときに操作等される。そして、入力装置2は、ユーザによる操作に対応する操作信号を、制御部3に供給する。

【0019】制御部3は、音声合成装置を構成する各ブロックを制御する。

【0020】文書解析部4は、制御部3の制御にしたがい、制御部3から供給されるテキスト(文書)を、単語辞書記憶部5や文法規則記憶部6を参照しながら言語処理する。即ち、単語辞書記憶部5には、各単語の品詞情報や、読み、アクセント等の情報が記述された単語辞書が記憶されており、また、文法規則記憶部6には、辞書記憶部5の単語辞書に記述された単語について、単語連鎖に関する制約等の文法規則が記憶されている。そして、文書解析部4は、この単語辞書および文法規則に基づいて、制御部3から供給されるテキストの形態素解析や構文解析等の解析を行い、その解析結果を、制御部3に供給する。

【0021】解析結果記憶部7は、文書解析部4において得られる解析結果を、制御部3を経由して記憶する。

【0022】音声記号生成部8は、アクセント規則記憶部9およびフレーズ規則記憶部10を参照しながら、制御部3から供給される、解析結果記憶部7に記憶された解析結果に基づいて、テキスト記憶部1に記憶されたテキストに対応する音声記号を生成する。

【0023】アクセント規則記憶部9は、アクセントを付与する規則(アクセント規則)を記憶しており、音声記号生成部8は、このアクセント規則にしたがって、音声記号に、アクセントを表す合成用タグ(例えば、前述の音声記号(1)に示した')を挿入する。フレーズ規則記憶部10は、フレーズを決定する際の規則(フレーズ規則)を記憶しており、音声記号生成部8は、このフレーズ規則にしたがって、音声記号に、フレーズを表す合成用タグ(例えば、前述の音声記号(1)に示した'')や('')を挿入する。

【0024】音声記号バッファ11は、制御部3の制御にしたがい、音声記号を一時記憶する。

【0025】規則音声合成部12は、制御部3から供給される音声記号にしたがい、音素片記憶部13を参照しながら、規則音声合成を行い、これにより、その音声記号に対応する合成音のデータ(合成音データ)を生成し、制御部3に供給する。即ち、音素片記憶部13には、例えば、C V(Consonant, Vowel)や、V C V、C V C等の形で音素片データが記憶されており、規則音声合成部12は、音声記号に基づいて、必要な音素片データを接続し、さらに、ポーズ、アクセント、イントネーシ

ョン等を適切に付加することで、合成音データを生成する。

【0026】音声波形バッファ14は、規則音声合成部12で生成された合成音データを一時記憶する。音声出力部15は、例えば、図示せぬD/A(Digital/Analog)変換器とスピーカ等で構成され、制御部3を経由して、音声波形バッファ14から供給される合成音データを、ディジタル信号からアナログ信号にD/A変換し、スピーカから出力する。

【0027】音声記号編集部16は、必要に応じて、制御部3を経由して、読み上げモード情報記憶部17を参照しながら、音声記号生成部8で生成された音声記号を編集する。読み上げモード情報記憶部17は、テキスト記憶部1に記憶されたテキストを合成音で読み上げる読み上げ方法についての各種の情報(読み上げモード情報)を記憶する。

【0028】図2は、図1の読み上げモード情報記憶部17に記憶される読み上げモード情報を示している。

【0029】図2の実施の形態において、読み上げモード情報「キャラクタ指定」は、テキストを読み上げる合成音のキャラクタを指定するときに「有り」とされ、指定しないときは「なし」とされる。なお、「キャラクタ指定」が「有り」の場合は、さらに、どのようなキャラクタ(例えば、男または女であるとか、歳が何歳くらいであるとか等)とするのかが、読み上げモード情報として設定される。

【0030】読み上げモード情報「吃りモード」は、合成音を吃了るものとするときに「ON」とされ、吃了のない合成音とするときは「OFF」とされる。

【0031】読み上げモード情報「反復回数」、「反復最大回数」、「反復最小回数」、「反復速度」、「反復最大速度」、「反復最小速度」、「反復加速度」、および「反復パワー変化」は、「吃りモード」が「ON」のときに有効になり、「吃りモード」が「OFF」のときは無効とされる(無視される)。

【0032】「反復回数」には、吃的回数、即ち、同一の音韻を反復する回数が設定される。なお、この回数は、乱数によって決定することが可能であり、この場合、「反復回数」は「自動」に設定される。

【0033】「反復最大回数」と「反復最小回数」には、音韻を反復する回数を乱数で決定する場合の、その回数の上限と下限が、それぞれ設定される。なお、「反復最大回数」および「反復最小回数」は、「反復回数」が「自動」のときのみ有効となり、「反復回数」に、数值が設定されているときは、無視される。

【0034】「反復速度」には、合成音の、音韻を反復する部分の発話速度が、例えば、1秒あたりのモーラ数(モーラ/秒)によって設定される。なお、この発話速度も、乱数によって決定することが可能であり、この場合、「発話速度」は「自動」に設定される。

【0035】「反復最大速度」と「反復最小速度」には、反復する音韻部分の発話速度を乱数で決定する場合の、その発話速度の上限と下限が、それぞれ設定される。なお、「反復最大速度」および「反復最小速度」は、「反復速度」が「自動」のときのみ有効となり、「反復速度」に、数値が設定されているときは、無視される。

【0036】「反復加速度」には、音韻を反復する部分の発話速度を変化させるときの、その変化量に関する情報が設定される。即ち、本実施の形態では、「反復加速度」には、音韻を反復する部分の最初と最後の発話速度が設定される。

【0037】「反復パワー変化」には、音韻を反復する部分のパワーを変化させるときの、その変化量に関する情報が設定される。即ち、本実施の形態では、「反復パワー変化」には、音韻を反復する部分の1モーラあたりのパワーの変化が設定される。

【0038】なお、読み上げモード情報は、入力装置2を操作することにより入力したり、する他、ネットワークを介し、図示せぬサーバからダウンロードして、読み上げモード情報記憶部17に記憶させることができる。さらに、読み上げモード情報記憶部17には、複数の読み上げモード情報を記憶させておき、その複数の読み上げモード情報の中から、読み上げのための音声合成に用いるものを、ユーザによる入力装置2の操作にしたがって選択するようにすることも可能である。

【0039】次に、図3のフローチャートを参照して、図1の音声合成装置による音声合成処理（ここでは、合成音によるテキストの読み上げ処理）について説明する。

【0040】ユーザが、入力装置2を操作し、テキスト記憶部1に記憶されたテキストの中から、読み上げの対象とするものを選択して、その読み上げを行うことを指示すると、入力装置2は、ステップS1において、読み上げを指示する指示信号を、制御部3に供給し、ステップS2に進む。

((' /) P3(' / ') . . .)

【0049】音声記号生成部8は、音声記号を生成すると、その音声記号を、制御部3を介して、音声記号バッファ11に供給して記憶させる。

【0050】そして、ステップS5に進み、制御部3は、読み上げモード情報記憶部17に、読み上げモード情報が記憶されているかどうかを判定し、記憶されていないと判定した場合、制御部3は、音声記号バッファ11に記憶された音声記号を、規則音声合成部12に供給して、ステップS6に進む。

【0051】ステップS6では、規則音声合成部12は、制御部3からの音声記号に含まれる音韻情報に対応する音韻に基づいて、必要な音素片データを、音素片記

【0041】ステップS2では、制御部3は、テキスト記憶部1から、読み上げ対象のテキストデータを読み出し、文書解析部4に供給して、ステップS3に進む。

【0042】ステップS3では、文書解析部4は、制御部3からのテキストデータについて、形態素解析や構文解析等を行い、その解析結果を、制御部3に供給する。制御部3は、文書解析部4からのテキストデータの解析結果を、解析結果記憶部7に供給して記憶させる。

【0043】ここで、読み上げ対象のテキストデータが、例えば、「これまで進められた研究は、大きな成果を・・・」であった場合の、文書解析部4による解析結果を、図4に示す。

【0044】解析結果には、図4に示すように、「形態素N○.」、「見出し」、「品詞」、「読み」、「活用」、「アクセント」等が含まれる。

【0045】「形態素N○.」は、読み上げ対象のテキストデータを構成する形態素に付されたユニークな番号を示す。「見出し」は、読み上げ対象のテキストデータ中の形態素を表す。「品詞」は、形態素の品詞を表す。「読み」は、形態素の音韻を表す。「活用」は、形態素の活用形を表す。「アクセント」は、形態素におけるアクセントの形を表し、平板型は0、頭高型は1、中高型はアクセントのおかれる拍の位置を表す。

【0046】図3に戻り、制御部3は、文書解析部4からのテキストデータの解析結果を、解析結果記憶部7に記憶させると、その旨を、音声記号生成部8に通知する。この場合、音声記号生成部8は、解析結果記憶部7に記憶された解析結果を、制御部3を介して読み出し、その解析結果に基づき、アクセントや、フレーズ、ポーズ等を表す合成用タグ、および音韻情報で構成される音声記号（列）を生成する。

【0047】ここで、テキストデータが「これまで進められた研究は、大きな成果を・・・」である場合には、音声記号生成部8では、例えば、前述の（1）で示したのと同様の、次のような音声記号（2）が生成される。

【0048】

/ ') . . .)

. . . (2)

憶部13から読み出し、ステップS7に進む。ステップS7では、規則音声合成部12は、音声記号に含まれるアクセント、フレーズ、およびアクセント句等を示す合成用タグに基づいて、合成音のピッチ周波数の変化パターンや、パワーの変化パターンを規定するイントネーションを作成し、ステップS8に進む。ステップS8では、規則音声合成部12は、ステップS6で得た音素片データを、ステップS7で作成したイントネーションや、音声記号に含まれるポーズ等を示す合成用タグに基づいて、波形を補正しながら接続していき、合成音データを生成する。規則音声合成部12は、合成音データを生成すると、その合成音データを、制御部3を介して、

音声波形バッファ14に供給して記憶させる。

【0052】そして、ステップS9に進み、制御部3は、音声波形バッファ14に記憶された合成音データを読み出し、音声出力部15に供給して、処理を終了する。これにより、音声出力部15では、ステップS2でテキスト記憶部1から読み出されたテキストデータに対応する合成音（そのテキストを読み上げる合成音）が 출력される。

【0053】一方、ステップS5において、読み上げモード情報記憶部17に、読み上げモード情報が記憶されていると判定された場合、制御部3は、音声記号バッファ11から音声記号を読み出し、その編集を要求する旨とともに、音声記号編集部16に供給して、ステップS10に進む。

【0054】ステップS10では、音声記号編集部16は、制御部3を介して、読み上げモード情報記憶部17から読み上げモード情報（図2）を読み出し、そのうちの「吃りモード」が「ON」になっているかどうかを判定する。ステップS10において、「吃りモード」が「ON」になっていないと判定された場合、ステップS11乃至S13をスキップして、ステップS14に進み、音声記号編集部16は、吃りに関係ない読み上げモード情報に基づいて、音声記号を編集し、ステップS15に進む。

【0055】即ち、例えば、音素片記憶部13に、音質の異なる音素片データが記憶されており、読み上げモード情報において、合成音の読み上げを行うキャラクタが設定されている場合には、音声記号編集部16は、そのキャラクタに対応する音質（あるいは、周波数特性）の

{(' /)P3(' /)...)

... (3)

【0060】なお、規則音声合成部12が、音韻を反復させることを指示する合成用タグ（以下、適宜、合成用反復タグという）を処理することができる場合には、ステップS10では、その合成用反復タグを、音声記号（2）に挿入するようにすることも可能である。即ち、例えば、合成用反復タグが、# rpt=#i#で定義され、その

{(¥ rpt=4¥ ' /)P3(' /)...)

... (4)

【0062】ここで、規則音声合成部12は、音声記号中に、合成用反復タグ# rpt=#i#が挿入されている場合、その合成用反復タグ# rpt=#i#の直後の音韻が、#i回だけ繰り返されるように、音素片データを接続する。

【0063】また、読み上げモード情報の「反復回数」が「自動」に設定されている場合には、音声記号編集部16は、音韻を反復させる回数を、乱数によって決定する。但し、この場合、乱数によって決定される回数としては、読み上げモード情報の「反復最大回数」に設定されている上限を越えず、かつ読み上げモード情報の「反

{(' /)P3(' /)...)

音素片データを用いることを指示する合成用タグを、音声記号に挿入する（例えば、音声記号の先頭に付加する）。

【0056】ステップS15では、音声記号編集部16は、編集後の音声記号を、制御部3を介して、音声記号バッファ11に供給して記憶させる。音声記号バッファ11において、編集後の音声記号が記憶されると、制御部3は、音声記号バッファ11に記憶された音声記号を、規則音声合成部12に供給して、ステップS6に進み、以下、上述の場合と同様の処理が行われる。

【0057】但し、例えば、上述のように、所定の音質の音素片データを用いることを指示する合成用タグが、音声記号に含まれている場合には、ステップS6では、その合成用タグによって指示されている音質の音素片データが、音素片記憶部13から読み出される。従って、この場合、音声出力部15から出力される合成音は、読み上げモード情報において設定されているキャラクタを想起させるような音質のものとなる。

【0058】一方、ステップS10において、「吃りモード」が「ON」になっていると判定された場合、ステップS11に進み、音声記号編集部16は、音声記号における文頭の音韻を、読み上げモード情報の「反復回数」に設定されている回数だけ反復させる。即ち、例えば、上述の音声記号（2）が得られている場合において、読み上げモード情報の「反復回数」が4に設定されている場合には、音声記号（2）は、その文頭の音韻「」を4回反復させた、次のような音声記号（3）に編集される。

【0059】

{(' /)...)

直後に配置される音韻を、#1回だけ反復させることを指示するものであるとすると、音声記号（2）の文頭の音韻を4回反復させる場合には、音声記号（2）は、次のような音声記号（4）に編集される。

【0061】

{(¥ rpt=4¥ ' /)...)

... (4)

「復最小回数」に設定されている下限を下回らない値が採用される。

【0064】さらに、ここでは、文頭の音韻を反復させようとしたが、文頭の音韻ではなく、または文頭の音韻とともに、例えば、ポーズがおわれる直後の音韻を反復させようすることも可能である。例えば、音声記号（2）について、文頭の音韻とともに、ポーズがおわれる直後の音韻を反復させた場合には、次のような音声記号（5）が得られる。

【0065】

{(' /)...)

【0066】ここで、音声記号(5)は、文頭の音韻「」を4回反復し、ポーズの直後の音韻「」を3回反復するものとなっている。

【0067】また、音声記号については、文頭やポーズの直後の音韻に限らず、任意の音韻を反復させることができある。どの音韻を反復させるかは、例えば、読み上げモード情報に設定することができる。

【0068】ステップS11において、音声記号における音韻を反復させた後は、ステップS12に進み、音声

$\{(\$spd=\text{標準}+2\$ \$spd=\text{標準}\$$
 $P3(\$spd=\text{標準}+2\$$

【0071】音声記号(6)には、発話速度を指示する合成用タグ(以下、適宜、合成用速度タグという) $\$spd=\#i\$$ が挿入されており、この合成用速度タグ $\$spd=\#i\$$ は、発話速度を、#i[モーラ/秒]とすることを意味する。規則音声合成部12は、音声記号中に、合成用速度タグ $\$spd=\#i\$$ が挿入されている場合、その合成用速度タグ $\$spd=\#i\$$ 以降に配置されている音韻の発話速度が、その後に新たな合成用速度タグ $\$spd=\#i\$$ が現れるまで、#i[モーラ/秒]となるように、音素片データを接続する。従って、音声記号(6)から生成される合成音については、文頭の4つの音韻「」は、標準+2[モーラ/秒]の発話速度で出力され、その後の、合成用速度タグ $\$spd=\text{標準}\$$ から $\$spd=\text{標準}+2\$$ までの音韻「」は、標準の発話速度で出力される。さら

に、ポーズ(P3)の後の合成用速度タグ $\$spd=\text{標準}+2\$$ の直後の3つの音韻「」は、標準+2[モーラ/秒]の発話速度で出力され、その後の合成用速度タグ $\$spd=\text{標準}\$$ に続く音韻「」は、標準の発話速度で出力される。

$\{(\$spd=\text{標準}+4\$ \$spd=\text{標準}\$$
 $P3(\$spd=\text{標準}+3\$$

【0076】音声記号(7)によれば、反復する音韻部分における音韻の反復回数に比例して、その反復する音韻部分の発話速度が速くなる合成音が生成される。

【0077】また、音声記号(6)や(7)では、反復する音韻部分において、その音韻を反復している間の発話速度は、一定となるが、その発話速度を変化させる編集を行うことも可能である。即ち、読み上げモード情報の「反復加速度」が設定されている場合には、音声記号

$\{(\$spd=2\text{to}8\$ \$spd=\text{標準}\$$
 $P3(\$spd=2\text{to}8\$$

【0079】ここで、音声記号(8)の合成用速度タグ $\$spd=2\text{to}8\$$ における2to8は、次の合成用速度タグが現れるまでの間の音韻の発話速度を、2モーラ/秒から8モーラ/秒に変化させていくことを表す。従って、この場

... (5)

記号編集部16は、音声記号について、その反復する音韻部分の発話速度を設定する編集を、読み上げモード情報を参照することで行う。

【0069】即ち、読み上げモード情報において、「反復速度」が、例えば、標準の発話速度+2[モーラ/秒]に設定されている場合には、例えば、上述の音声記号(5)は、次のような音声記号(6)に編集される。

【0070】

/)
 $\$spd=\text{標準}\$ / ') \dots$
 ... (6)

【0072】なお、音声記号(6)では、反復する音韻部分「」および「」が、他の音韻部分よりも2[モーラ/秒]だけ、発話速度が速くなるが、これは、反復する音韻部分の発話速度を、他の音韻部分よりも速くした方が、自然な吃りの合成音が得られるからである。

【0073】また、読み上げモード情報の「反復速度」が「自動」に設定されている場合には、音声記号編集部16は、反復する音韻部分の発話速度を、乱数によって決定する。但し、この場合、乱数によって決定される発話速度としては、読み上げモード情報の「反復最大速度」に設定されている上限を越えず、かつ読み上げモード情報の「反復最小速度」に設定されている下限を下らない値が採用される。

【0074】さらに、反復する音韻部分の発話速度は、その音韻の反復回数によって設定することが可能である。この場合、例えば、上述の音声記号(5)は、次のような音声記号(7)に編集される。

【0075】

/)
 $\$spd=\text{標準}\$ / ') \dots$
 ... (7)

編集部16は、その「反復加速度」の設定にしたがい、反復する音韻部分において、その音韻を反復している間の発話速度が変化するよう、音声記号を編集する。具体的には、例えば、読み上げモード情報の「反復加速度」が、図2に示したように、「2モーラ/秒→8モーラ/秒」に設定されている場合には、音声記号(5)は、次のような音声記号(8)に編集される。

【0078】

/)
 $\$spd=\text{標準}\$ / ') \dots$
 ... (8)

合、規則音声合成部12では、反復される音韻部分において、そのように発話速度が変化するよう、音素片データが接続される。

【0080】ステップS12において、上述のように、

音声記号に、合成用速度タグを挿入する編集を行った後は、ステップS13に進み、音声記号編集部16は、音声記号について、その反復する音韻部分のパワーを設定する編集を、読み上げモード情報を参照することで行う。

【0081】即ち、読み上げモード情報において、「反

```
((#pow=+4#$spd=2to8#$spd=標準#$pow=標準# / )  
P3(#pow=+4#$spd=2to8#$spd=標準#$pow=標準# / )...)  
... (9)
```

【0083】音声記号(9)には、パワーを指示する合成用タグ(以下、適宜、合成用パワータグという)#\$pow=#i#が挿入されており、この合成用パワータグ#\$pow=#i#は、パワーを、#i [dB]とすることを意味する。但し、合成用パワータグ#\$pow=±i#は、パワーを、1モーラあたり、± i [dB]ずつ変化させることを表す。規則音声合成部12は、音声記号中に、合成用パワータグ#\$pow#が挿入されている場合、その合成用速度タグ#\$pow#以降に配置されている音韻のパワーを、その合成用速度タグ#\$pow#にしたがって制御する。従って、音声記号(9)から生成される合成音については、文頭の4つの音韻「」は、1モーラあたり4 dBずつ増加するパワーで出力され、その後の、合成用パワータグ#\$pow=標準#から#\$pow=+4#までの音韻「」は、標準のパワーで出力される。さらに、ポーズ(P3)の後の合成用速度タグ#\$pow=+4#の直後の3つの音韻「」は、再び、1モーラあたり4 dBずつ増加するパワーで出力され、その後の合成用速度タグ#\$pow=標準#に続く音韻「」は、標準のパワーで出力される。

【0084】なお、その他、例えば、反復される音韻部分の直後の音韻のパワーが大となるような編集を行うことも可能である。

【0085】ステップS13において、上述のように、音声記号に、合成用パワータグを挿入する編集を行った後は、ステップS14に進み、以下、上述の場合と同様の処理が行われ、音声出力部15において合成音が出力される。

【0086】従って、この場合、音声出力部15から出力される合成音は、所定の音韻が、読み上げモード情報にしたがって反復回数、発話速度、パワーが制御されたもの、即ち、自然な吃りのある合成音となる。

【0087】以上のように、音声記号を、読み上げモード情報にしたがって編集することにより、音声記号における所定の音韻を反復させるようにしたので、容易に、吃りのある合成音を生成することができる。

【0088】さらに、音声記号について、反復させる音韻部分の発話速度やパワーを制御する編集も行うようにしたので、自然な吃りのある合成音を生成することができる。

【0089】次に、上述した一連の処理は、ハードウェアにより行うこともできるし、ソフトウェアにより行う

復パワー変化」が、例えば、図2に示したように、「+4 dB/モーラ」に設定されている場合には、例えば、上述の音声記号(8)は、次のような音声記号(9)に編集される。

【0082】

ことでもできる。一連の処理をソフトウェアによって行う場合には、そのソフトウェアを構成するプログラムが、汎用のコンピュータ等にインストールされる。

【0090】そこで、図5は、上述した一連の処理を実行するプログラムがインストールされるコンピュータの一実施の形態の構成例を示している。

【0091】プログラムは、コンピュータに内蔵されている記録媒体としてのハードディスク105やROM103に予め記録しておくことができる。

【0092】あるいはまた、プログラムは、フレキシブルディスク、CD-ROM(Compact Disc Read Only Memory), MO(Magneto optical)ディスク、DVD(Digital Versatile Disc)、磁気ディスク、半導体メモリなどのリムーバブル記録媒体111に、一時的あるいは永続的に格納(記録)しておくことができる。このようなリムーバブル記録媒体111は、いわゆるパッケージソフトウェアとして提供することができる。

【0093】なお、プログラムは、上述したようなリムーバブル記録媒体111からコンピュータにインストールする他、ダウンロードサイトから、ディジタル衛星放送用の人工衛星を介して、コンピュータに無線で転送したり、LAN(Local Area Network)、インターネットといったネットワークを介して、コンピュータに有線で転送し、コンピュータでは、そのようにして転送されてくるプログラムを、通信部108で受信し、内蔵するハードディスク105にインストールすることができる。

【0094】コンピュータは、CPU(Central Processing Unit)102を内蔵している。CPU102には、バス101を介して、入出力インターフェース110が接続されており、CPU102は、入出力インターフェース110を介して、ユーザによって、キーボードや、マウス、マイク等で構成される入力部107が操作等されることにより指令が入力されると、それにしたがって、ROM(Read Only Memory)103に格納されているプログラムを実行する。あるいは、また、CPU102は、ハードディスク105に格納されているプログラム、衛星若しくはネットワークから転送され、通信部108で受信されてハードディスク105にインストールされたプログラム、またはドライブ109に装着されたリムーバブル記録媒体111から読み出されてハードディスク105にインストールされたプログラムを、RAM(Random Access Memor

y) 104にロードして実行する。これにより、CPU102は、上述したフローチャートにしたがった処理、あるいは上述したブロック図の構成により行われる処理を行う。そして、CPU102は、その処理結果を、必要に応じて、例えば、入出力インターフェース110を介して、LCD(Liquid Crystal Display)やスピーカ等で構成される出力部106から出力、あるいは、通信部108から送信、さらには、ハードディスク105に記録等させる。

【0095】ここで、本明細書において、コンピュータに各種の処理を行わせるためのプログラムを記述する処理ステップは、必ずしもフローチャートとして記載された順序に沿って時系列に処理する必要はなく、並列的あるいは個別に実行される処理（例えば、並列処理あるいはオブジェクトによる処理）も含むものである。

【0096】また、プログラムは、1のコンピュータにより処理されるものであっても良いし、複数のコンピュータによって分散処理されるものであっても良い。さらに、プログラムは、遠方のコンピュータに転送されて実行されるものであっても良い。

【0097】以上、本発明の音声合成装置について説明したが、この音声合成装置は、現実世界における実際のロボットや、表示装置に表示される仮想的なロボット（キャラクタ）、対話システムその他に広く適用することができる。例えば、ロボットに適用した場合には、ロボットが、吃りのある合成音による発話をを行うことから、ロボットに個性を持たせることができ、そのエンタテイメント性を向上させることができる。

【0098】なお、本実施の形態では、音声記号編集部16を、規則音声合成部12とは別に構成するようにしたが、音声記号編集部16は、規則音声合成部12に内蔵させる形で構成することも可能である。

【0099】また、本実施の形態では、音声記号の編集を、読み上げモード情報にしたがって行うようにしたが、音声記号の編集は、読み上げモード情報に対応する編集用タグを、音声記号に、あらかじめ挿入しておき、その編集用タグにしたがって行うようにすることも可能である。

【0100】即ち、例えば、音声記号の編集において音韻を反復させることを指示するタグ（以下、適宜、編集用反復タグという）#vch=Domori¥を、編集用タグの1つとして定義しておき、音声記号編集部16には、このような編集用タグにしたがって、音声記号の編集を行わせることができる。

【0101】この場合、例えば、音声記号が「{(' / #vch=Domori¥ /)P3(' / ')...}」であったときには、音声記号編集部16では、編集用反復タグ#vch=Domori¥にしたがい、「{(' / ... /)P3(' / ')...}」のように、編

集用反復タグ#vch=Domori¥の直後の音韻「」を所定回数だけ反復させるように編集が行われる。

【0102】なお、この場合、読み上げモード情報は不要となる。但し、読み上げモード情報と、編集用タグとは、併用することが可能である。

【0103】また、編集用タグは、音声記号ではなく、その音声記号を生成する前のテキストデータに挿入しておくことが可能である。なお、音韻情報を有する記号（列）を、音声情報というものとすると、音声記号や、その音声記号を生成する前のテキストデータは、いずれも音声情報であり、上述した編集は、音声情報を対象として行なうことができる。即ち、上述の編集は、音声記号ではなく、テキストデータを対象に行なうことも可能である。

【0104】また、本実施の形態では、合成音を規則音声合成によって生成するようにしたが、合成音は、規則音声合成以外の音声合成によって生成することも可能である。

【0105】

【発明の効果】以上のごく、本発明の音声合成装置および音声合成方法、並びにプログラムによれば、音声情報が、所定の規則にしたがって編集され、音声情報における所定の音韻を反復させた状態の音声情報に変換される。そして、その編集された音声情報にしたがい、合成音が生成される。従って、吃りのある合成音を、容易に生成することができる。

【図面の簡単な説明】

【図1】本発明を適用した音声合成装置の一実施の形態の構成例を示すブロック図である。

【図2】読み上げモード情報を示す図である。

【図3】音声合成装置の処理を説明するフローチャートである。

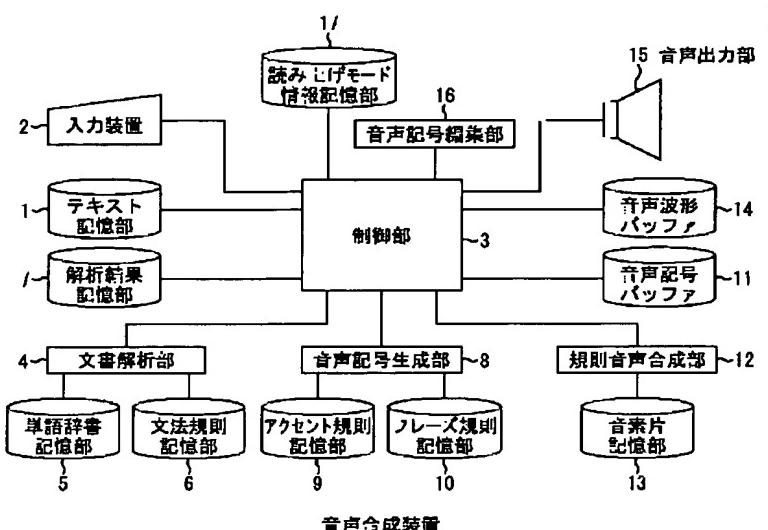
【図4】文書解析部4の処理結果を示す図である。

【図5】本発明を適用したコンピュータの一実施の形態の構成例を示すブロック図である。

【符号の説明】

- 1 テキスト記憶部,
- 2 入力装置,
- 3 制御部,
- 4 文書解析部,
- 5 単語辞書記憶部,
- 6 文法規則記憶部,
- 7 解析結果記憶部,
- 8 音声記号生成部,
- 9 アクセント規則記憶部,
- 10 フレーズ規則記憶部,
- 11 音声記号バッファ,
- 12 規則音声合成部,
- 13 音素片記憶部,
- 14 音声波形バッファ,
- 15 音声出力部,
- 16 音声記号編集部,
- 17 読み上げモード情報記憶部,
- 101 バス,
- 102 CPU,
- 103 ROM,
- 104 RAM,
- 105 ハードディスク,
- 106 出力部,
- 107 入力部,
- 108 通信部,
- 109 ドライブ,
- 110 入出力インターフェース,
- 111 リムーバブル記録媒体

【図1】



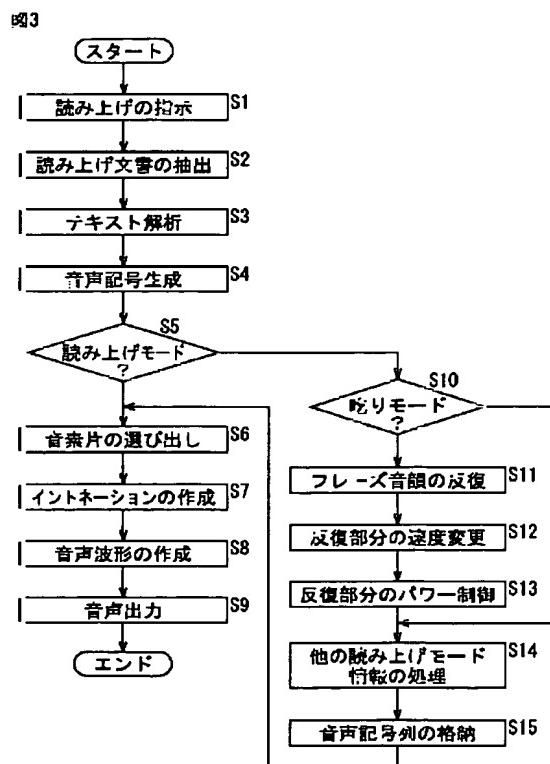
【図2】

図2

...	...
キャラクタ指定	有り
吃りモード	ON
反復回数	自動
反復最大回数	7
反復最小回数	3
反復速度	自動
反復最大速度	8モーラ/秒
反復最小速度	2モーラ/秒
反復加速度	2モーラ/秒 → 8モーラ/秒へ
反復パワー変化	+4db/モーラ
...	...

読み上げモード情報

【図3】



【図4】

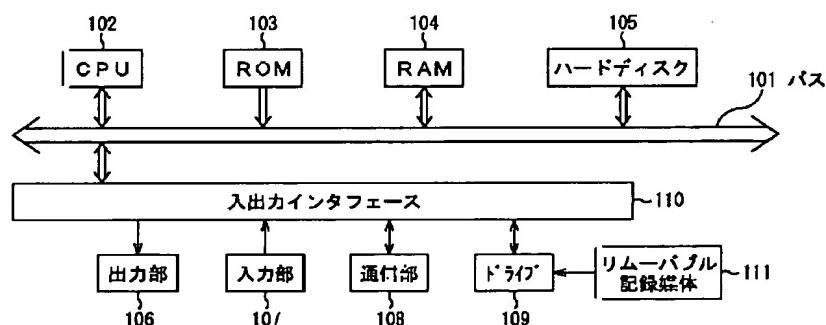
図4

形態素No.	0	1	2	3	4	5
見出し	これ	まで	進め	られ	た	研究
品詞	代名詞	助詞	-段動詞	助動詞	助動詞	よ変名詞
読み	コレ	メテ	スルメ	ル	タ	ケキョウ
活用	未然	連用	連体
アクセント	0	1	0	0	0	0
....

解析結果

【図5】

図5



コンピュータ

フロントページの続き

(72)発明者 浅野 康治
東京都品川区北品川6丁目7番35号 ソニ
一株式会社内

(72)発明者 富谷 真一
東京都品川区北品川6丁目7番35号 ソニ
一株式会社内

(72)発明者 藤田 八重子
東京都品川区北品川6丁目7番35号 ソニ
一株式会社内
F ターム(参考) 5D045 AA01 AA07 AB11